IN THE CLAIMS:

The following is a complete list of the claims now pending. This listing replaces all earlier versions and listings of the claims.

Claim 1 (currently amended): ~~Apparatus~~ An apparatus for processing image data and sound data, comprising:

an image processor for processing image data recorded by at least one camera showing the movements of a plurality of people to track each person in three dimensions;

a sound processor for processing sound data to determine the direction of arrival of the sound;

a speaker identifier for determining which of the people is speaking based on the result of the processing performed by [[the]] said image processor and the result of the processing performed by [[the]] said sound processor; and

a voice recognition processor for processing the received sound data to generate text data therefrom in dependence upon the result of the processing performed by [[the]] said speaker identifier.

Claim 2 (currently amended): ~~Apparatus~~ An apparatus according to claim 1, wherein [[the]] said voice recognition processor includes a [[store]] storage unit for storing respective voice recognition parameters for each of the plurality of people, and a selection

processor for selecting the voice recognition parameters to be used to process the sound data in dependence upon the person determined to be speaking by the speaker identifier.

Claim 3 (currently amended): ~~Apparatus~~ An apparatus according to claim 1, wherein [[the]] said image processor is arranged to track each person by processing the image data using camera calibration data defining the position and orientation of each camera from which image data is processed.

Claim 4 (currently amended): ~~Apparatus~~ An apparatus according to claim 1, wherein [[the]] said image processor is arranged to track each person by tracking each person's head.

Claim 5 (currently amended): ~~Apparatus~~ An apparatus according to claim 1, wherein [[the]] said image processor is arranged to process the image data to determine where at least each person who is speaking is looking.

Claim 6 (currently amended): ~~Apparatus~~ An apparatus according to claim 1, wherein [[the]] said speaker identifier is arranged to identify a person who is speaking in a given frame of the received image data using the results of the processing performed by [[the]] said image processor and [[the]] said sound processor for at least one other frame if

the speaker cannot be identified using the results of the processing performed by [[the]] said image processor and [[the]] said sound processor for the given frame.

Claim 7 (currently amended): ~~Apparatus~~ An apparatus according to claim 1, further comprising a database for storing at least some of the received image data, the sound data, the text data produced by [[the]] said voice recognition processor and viewing data defining where at least each person who is speaking is looking, [[the]] said database being arranged to store the data such that corresponding text data and viewing data are associated with each other and with the corresponding image data and sound data.

Claim 8 (currently amended): ~~Apparatus~~ An apparatus according to claim 7, further comprising a data compressor for compressing the image data and the sound data for storage in [[the]] said database.

Claim 9 (currently amended): ~~Apparatus~~ An apparatus according to claim 8, wherein [[the]] said data compressor comprises a data encoder for encoding the image data and the sound data as MPEG data.

Claim 10 (currently amended): ~~Apparatus~~ An apparatus according to claim 7, further comprising a gaze data generator for generating data defining, for a predetermined period, the proportion of time spent by a given person looking at each of the

other people during the predetermined period, and wherein [[the]] said database is arranged to store the data so that it is associated with the corresponding image data, sound data, text data and viewing data.

Claim 11 (currently amended): ~~Apparatus~~ An apparatus according to claim 10, wherein the predetermined period comprises a period during which the given person was talking.

B1 Cont

Claim 12 (currently amended): ~~Apparatus~~ An apparatus for processing image data and sound data, comprising:

an image processor for processing image data recorded by at least one camera showing the movements of a plurality of people to track each person in three dimensions;

a sound processor for processing sound data to determine the direction of arrival of the sound; and

a speaker identifier for determining which of the people is speaking based on the result of the processing performed by [[the]] said image processor and the result of the processing performed by [[the]] said sound processor.

Claim 13 (currently amended): ~~Apparatus~~ An apparatus according to claim 12, wherein [[the]] said image processor is arranged to track each person by processing the

image data using camera calibration data defining the position and orientation of each camera from which image data is processed.

Claim 14 (currently amended): ~~Apparatus~~ An apparatus according to claim 12, wherein [[the]] said image processor is arranged to track each person by tracking each person's head.

Claim 15 (currently amended): ~~Apparatus~~ An apparatus according to claim 12, wherein [[the]] said image processor is arranged to process the image data to determine where at least each person who is speaking is looking.

Claim 16 (currently amended): ~~Apparatus~~ An apparatus according to claim 12, wherein [[the]] said speaker identifier is arranged to identify a person who is speaking in a given frame of the received image data using the results of the processing performed by [[the]] said image processor and [[the]] said sound processor for at least one other frame if the speaker cannot be identified using the results of the processing performed by [[the]] said image processor and [[the]] said sound processor for the given frame.

Claim 17 (currently amended): A method of processing image data and sound data, comprising:

an image processing step, of ~~comprising~~ processing image data recorded by at least one camera showing the movements of a plurality of people to track each person in three dimensions;

a sound processing step, of ~~comprising~~ processing sound data to determine the direction of arrival of the sound;

a speaker identification step, of ~~comprising~~ determining which of the people is speaking based on the result of the processing performed in [[the]] said image processing step and the result of the processing performed in [[the]] said sound processing step; and

B1 cont'd

a voice recognition processing step, of ~~comprising~~ processing the received sound data to generate text data therefrom in dependence upon the result of the processing performed in [[the]] said speaker identification step.

Claim 18 (currently amended): A method according to claim 17, wherein, [[the]] said voice recognition processing step includes selecting, from stored respective voice recognition parameters for each of the plurality of people, [[the]] voice recognition parameters to be used to process the sound data in dependence upon the person determined to be speaking in [[the]] said speaker identification step.

Claim 19 (currently amended): A method according to claim 17, wherein, ~~in the~~ said image processing step, ~~each person is tracked~~ includes tracking each person by

processing the image data using camera calibration data defining the position and orientation of each camera from which image data is processed.

Claim 20 (currently amended): A method according to claim 17, wherein, ~~in the~~ said image processing step~~, each person is tracked~~ includes tracking each person by tracking the person's head.

Claim 21 (currently amended): A method according to claim 17, wherein, ~~in the~~ said image processing step~~, the image data is processed~~ includes processing the image data to determine where at least each person who is speaking is looking.

Claim 22 (currently amended): A method according to claim 17, wherein, ~~in the~~ said speaker identification step[[,]] includes identifying a person who is speaking in a given frame of the received image data ~~is identified~~ using the results of the processing performed in [[the]] said image processing step and [[the]] said sound processing step for at least one other frame if the speaker cannot be identified using the results of the processing performed in [[the]] said image processing step and [[the]] said sound processing step for the given frame.

Claim 23 (currently amended): A method according to claim 17, further comprising [[the]] a signal generating step, of generating a signal conveying the text data generated in [[the]] said voice recognition processing step.

Claim 24 (currently amended): A method according to claim 17, further comprising [[the]] a received image data storage step, of storing in a database at least some of the received image data, the sound data, the text data produced in [[the]] said voice recognition processing step and viewing data defining where at least each person who is speaking is looking, the data being stored in the database such that corresponding text data and viewing data are associated with each other and with the corresponding image data and sound data.

Claim 25 (original): A method according to claim 24, wherein the image data and the sound data are stored in the database in compressed form.

Claim 26 (original): A method according to claim 25, wherein the image data and the sound data are stored as MPEG data.

Claim 27 (currently amended): A method according to claim 24, further comprising:

~~the steps~~ a data defining generation step, of generating data defining, for a predetermined period, the proportion of time spent by a given person looking at each of the other people during the predetermined period[[,]]; and

a data storage step, of storing the data in the database so that it is associated with the corresponding image data, sound data, text data and viewing data.

Claim 28 (original): A method according to claim 27, wherein the predetermined period comprises a period during which the given person was talking.

Claim 29 (currently amended): A method according to claim 24, further comprising [[the]] a generating step, of generating a signal conveying the database with data therein.

Claim 30 (currently amended): A method according to claim 29, further comprising [[the]] a recording step, of recording the signal either directly or indirectly to generate a recording thereof.

Claim 31 (currently amended): A method of processing image data and sound data, comprising:

an image processing step, of ~~comprising~~ processing image data recorded by at least one camera showing the movements of a plurality of people to track each person in three dimensions;

a sound processing step, of ~~comprising~~ processing sound data to determine the direction of arrival of the sound; and

a speaker identification step, of ~~comprising~~ determining which of the people is speaking based on the result of the processing performed in [[the]] said image processing step and the result of the processing performed in [[the]] said sound processing step.

Claim 32 (currently amended): A method according to claim 31, wherein ~~in the~~ said image processing step, ~~each person is tracked~~ includes tracking each person by processing the image data using camera calibration data defining the position and orientation of each camera from which image data is processed.

Claim 33 (currently amended): A method according to claim 31, wherein ~~in the~~ said image processing step, ~~each person is tracked by~~ includes tracking each person by tracking the person's head.

Claim 34 (currently amended): A method according to claim 31, wherein ~~in the~~ said image processing step~~, the image data is processed~~ includes processing the image data to determine where at least each person who is speaking is looking.

Claim 35 (currently amended): A method according to claim 31, wherein ~~in the~~ speaker identification step[[,]] includes identifying a person who is speaking in a given frame of the received image data ~~is identified~~ using the results of the processing performed in [[the]] said image processing step and [[the]] said sound processing step for at least one other frame if the speaker cannot be identified using the results of the processing performed in [[the]] said image processing step and [[the]] said sound processing step for the given frame.

Claim 36 (currently amended): A method according to claim 31, further comprising the step of generating a signal conveying the identity of the speaker identified in [[the]] said speaker identification step.

Claim 37 (currently amended): A storage device storing computer program instructions for programming ~~causing~~ a programmable processing apparatus to become configured as an apparatus as set out in ~~at least~~ any one of claims 1, [[and]] 12, 87 and 88.

Claim 38 (currently amended): A storage device storing computer program instructions for programming ~~causing~~ a programmable processing apparatus to become operable to perform a method as set out in ~~at least~~ any one of claims 17, [[and]] 31, 89 and 90.

Claim 39 (currently amended): A signal conveying computer program instructions for programming ~~causing~~ a programmable processing apparatus to become configured as an apparatus as set out in ~~at least~~ any one of claims 1, [[and]] 12, 87 and 88.

Claim 40 (currently amended): A signal conveying computer program instructions for programming ~~causing~~ a programmable processing apparatus to become operable to perform a method as set out in ~~at least~~ any one of claims 17, [[and]] 31, 89 and 90.

Claim 41 (currently amended): ~~Apparatus~~ An apparatus for processing image data and sound data, comprising:

image processing means for processing image data recorded by at least one camera showing the movements of a plurality of people to track each person in three dimensions;

sound processing means for processing sound data to determine the direction of arrival of the sound;

speaker identification means for determining which of the people is speaking based on the result of the processing performed by [[the]] said image processing means and the result of the processing performed by [[the]] said sound processing means; and

voice recognition processing means for processing the received sound data to generate text data therefrom in dependence upon the result of the processing performed by [[the]] said speaker identification means.

Claim 42 (currently amended): ~~Apparatus~~ An apparatus for processing image data and sound data, comprising:

image processing means for processing image data recorded by at least one camera showing the movements of a plurality of people to track each person in three dimensions;

sound processing means for processing sound data to determine the direction of arrival of the sound; and

speaker identification means for determining which of the people is speaking based on the result of the processing performed by [[the]] said image processing means and the result of the processing performed by [[the]] said sound processing means.

Claims 43-86 (canceled)

Claim 87 (new): An apparatus for processing image data and sound data, comprising:

an image processor operable to process image data recorded by at least one camera showing the movements of a plurality of people to track each person;

a sound processor operable to process sound data to determine the direction of arrival of the sound; and

a speaker identifier operable to determine which of the people is speaking based on the result of the processing performed by said image processor and the result of the processing performed by said sound processor.

Claim 88 (new): An apparatus for processing image data and sound data, comprising:

an image processor operable to process image data recorded by at least one camera showing the movements of a plurality of people in three dimensions;

a sound processor operable to process sound data to determine the direction of arrival of the sound; and

a speaker identifier operable to determine which of the people is speaking based on the result of the processing performed by said image processor and the result of the processing performed by said sound processor.

Claim 89 (new): A method of processing image data and sound data, comprising:

an image processing step, of processing image data recorded by at least one camera showing the movements of a plurality of people to track each person;

a sound processing step, of processing sound data to determine the direction of arrival of the sound; and

a speaker identification step, of determining which of the people is speaking based on the result of the processing performed in said image processing step and the result of the processing performed in said sound processing step.

Claim 90 (new): A method of processing image data and sound data, comprising:

an image processing step, of processing image data recorded by at least one camera showing the movements of a plurality of people in three dimensions;

a sound processing step, of processing sound data to determine the direction of arrival of the sound; and

a speaker identification step, of determining which of the people is speaking based on the result of the processing performed in said image processing step and the result of the processing performed in said sound processing step.

Claim 91 (new): An apparatus for processing image data and sound data, comprising:

image processing means for processing image data recorded by at least one camera showing the movements of a plurality of people to track each person;

sound processing means for processing sound data to determine the direction of arrival of the sound; and

speaker identification means for determining which of the people is speaking based on the result of the processing performed by said image processing means and the result of the processing performed by said sound processing means.

Claim 92 (new): An apparatus for processing image data and sound data, comprising:

image processing means for processing image data recorded by at least one camera showing the movements of a plurality of people in three dimensions;

sound processing means for processing sound data to determine the direction of arrival of the sound; and

speaker identification means for determining which of the people is speaking based on the result of the processing performed by said image processing means and the result of the processing performed by said sound processing means.